Editors' Suggestion

Pulse instabilities can shape virus-immune coevolution

David A. Kessler
Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel

Herbert Levine
Center for Theoretical Biological Physics and Departments of Physics and Bioengineering, Northeastern University, Boston, Massachusetts 02215, USA

(Received 30 October 2023; accepted 9 August 2024; published 16 September 2024)

Adaptive immune systems engage in an arms race with evolving viruses, trying to generate new responses to viral strains that continually move away from the set of genetically varying strains that have already elicited a functional immune response. It has been argued that this dynamical process can lead to a propagating pulse of an ever-changing viral population and concomitant immune response. Here, we introduce a new stochastic model of viral-host coevolution, taking into account finite-sized host populations and varying processes of immune "forgetting". Using both stochastic and deterministic calculations, we show that there is indeed a possible pulse solution, but for a large host population size and for finite memory capacity, the pulse becomes unstable to the generation of new infections in its wake. This instability leads to an extended endemic infection pattern, demonstrating that the population-level behavior of virus infections can exhibit a wider range of behavior than had been previously realized.

DOI: 10.1103/PhysRevResearch.6.033300

I. INTRODUCTION

Viruses form a convenient forum for studying Darwinian evolution [1]. They reproduce quickly, giving rise to large populations, and are therefore more deterministic in their dynamics than populations of macroscopic living creatures [2]. Although their reproduction is asexual, nevertheless they can even exhibit rudimentary forms of genetic recombination when different viral strains with multicompartmented genomes invade a common cell [3].

A powerful conceptual approach for studying viral evolutionary dynamics involves projecting the dynamics onto a low-dimensional fitness space [4–6]. This approach led to the discovery of the possibility of a propagating solitary wave of viral strains on a linear fitness landscape, propelled by the more rapid growth of the advancing edge as compared to the out-competed variants at the decaying tail. Interestingly, determining the velocity of this wave turned out to be a challenging statistical physics problem, as the naive continuum partial differential equation (PDE) governing the viral fitness density exhibits a finite time singularity [5] and hence necessitates a more nuanced treatment of this limit [6,7]. More specifically, a modified PDE approach [5,6,8] led to the prediction of logarithmic scaling of the advancing speed with viral population size, a result in agreement with simulations of stochastic versions of the model and with the exact result for a specific model variant [9].

Most recently, a series of papers [10–12] have generalized this framework to study the coevolution of the viral population with the immune system. Here, individual hosts generate immune responses that target the virus even as the virus tries to mutate away from immune recognition. Of particular importance in this class of models is the extent to which the immune response, mediated for example via antibody recognition of the viral particle, decays as a function of mutational distance [13]. One possible simple assumption is that recognition falls off as an exponential of the Hamming distance between D-dimensional viral genomes; one can also imagine nonsymmetric falloffs [10], but the reasoning behind this assumption is not obvious. As pointed out in these papers, the immune response to previous infections can create an effective gradient in viral fitness; the further away from the bulk of the population, the weaker the immune suppression and hence the faster the growth. This means that the propagation of a single pulse along some direction in this D-dimensional space is in the same class [14,15] as the aforementioned pure viral evolution problem. The fact that such a pulse solution is localized in the direction transverse to the propagation [11] makes this type of one-dimensional solution semiquantitatively relevant even for an arbitrary number of phenotypic dimensions. Thus, predictions from this modeling framework can be very informative regarding general patterns of viral dynamics.

In this paper, we focus our attention on the one-dimensional case and study a generalized version of the model of Ref. [11]. Our study considers the stochastic dynamics together with the modified PDE approach wherein the leading effects of stochasticity (here, demographic noise) are taken into account by a small-population cutoff of the interaction dynamics. In the deterministic version, we show that a stable single pulse solution exists for smaller host populations but

that it becomes (nonlinearly) unstable above some critical population size. This instability leads, immediately beyond the critical size, to a patterned tail and eventually, at even larger sizes, to a featureless tail. Both of these are forms of endemic infections. These expanding tails cause the front to slow down with time as they exhaust the number of available hosts for new variants to infect. Stochastic dynamics blurs the distinction between the patterned and unpatterned tail, but otherwise exhibits a similar transition from a stable pulse to an endemic state exhibiting a slowly widening and eventually stationary band of active variants. Thus, the long-time behavior is often a nonspreading statistically steady infection pattern. This new type of behavior for viral-host coevolutionary dynamics, possibly relevant both for human respiratory diseases and for the infection of bacteria by phages, is the major result of this paper.

II. STOCHASTIC MODEL

A. Formulation

The basic process we propose to study involves an evolving virus that infects individual hosts in a population. Hosts respond to infection by launching an immune response, best thought of as selecting antibodies that effectively neutralize viral particles emitted by infected individuals. Of specific interest is the fact that such antibodies will also, to some extent, neutralize mutant strains. This extent depends on exactly how far the mutant has strayed from the original.

The actual problem of determining the expected falloff of immune effectiveness is rather complex and undoubtedly context dependent. However, interesting insights into the behavior of viral immune coevolution have been obtained by making simplified assumptions about this feature of "shape space" [16]. The simplest assumption is that there is a monotonic mapping of immune effectiveness falloff as a function of how many relevant mutations have been accumulated by a particular viral strain [11,13]. Using this approach, we can define a cross-reactivity function $g(|\vec{x} - \vec{x}'|)$, where the fall-off function g is the ratio of effectiveness at a distance $d = |\vec{x} - \vec{x}'|$) to that at zero distance; g obviously approaches unity at small distances and vanishes at large distances. This assumption is reminiscent of that used in ecological models of competing species [17,18], where niche selectivity is assumed to give rise to a nonlocal but decaying interaction between specialists optimized to grow in specific environments. In what follows, we will assume that g takes the form of an exponential of width w. As noted above, we here consider a discrete one-dimensional shape space, with position (i.e., strain) labeled by the index i.

Our model assumes that there are a total of N_h individual hosts, each of which can be infected by at most one viral strain at a time. The number of infected hosts is $N_I(t)$, which obviously cannot exceed N_h. The possible events at some given moment are the recovery of one of the infected hosts (at total rate rN_I) or the release by one of the hosts of a new virion at rate rN_IR_0. Here the base reproduction rate R_0 is the mean total number of "effective" virions, namely virions that are directly able to infect new individuals, released by a single infected host during its duration of being infected before recovery. Given the total rate of events, we choose the time step according to the standard Gillespie algorithm and then pick which event occurs with the relative probabilities $1/(1 + R_0)$ for clearance and $R_0/(1 + R_0)$ for virion release, and for both event types the host in question is chosen at random from the infected pool. Clearance is effected by removing the chosen host from the infected list, reducing N_I by one, and updating its immune memory (see below). The release of a virion is accompanied by its attempt to infect a new host. We allow this released virion to be slightly modified from the strain that caused the infection in question; specifically, we allow its index i in shape space to mutate either up or down by one with equal probabilities given by $\mu/2$. If the randomly chosen new host is currently uninfected, then infection will occur with a probability $p_{\rm inf}$ that depends on the degree to which that host has immunity to that particular virion strain.

The model is completed by describing how the immune state of an individual host is updated and how that state determines the infection probability. We will assume that the host carries multiple immune "memories", corresponding to its previous infection history. The immune memory is realized in the matrix $mem(j, i)$, storing the memory of host j having been infected by viral strain i. There is no rule that memories need to be distinct, as hosts can have been multiply infected in the past by the same strain, with each exposure bolstering the cell's immunity to that strain. In our update scheme, every time a host recovers, it stores a memory of the strain from which it has just recovered. In the usual case where the number of existing memories for that host is equal to the memory capacity, this new memory overwrites a randomly chosen previous memory. Finally, we have the explicit formula for the probability of infecting host j by viral strain k

$$p_{\rm inf}(j, k) = \prod_{i=1}^{M} (1 - p_0 g(|mem(j, i) - k|)). \qquad (1)$$

Finally, we choose $p_0 = (1 - p_{00}^{1/M})$, which therefore gives a bare infection rate of p_{00} for any hosts all of whose memories are equal to that of the strain with which it is being challenged. A brief description of the variables and parameters in our model is presented in Table I and a schematic is presented in Fig. 1.

There are obviously many parameters in this model and it would be impossible to exhaustively vary all of them. As our interest here is in revealing the types of behavior that can occur, we will initially stick to the choices $p_{00} = 0.015$, $M = 6$, giving $p_0 \approx 0.50$. We will consider two values for the mutation rate, a high value (0.5) which creates relatively smooth profiles and hence allows comparison to our deterministic PDE approach, and a perhaps more biologically reasonable value (0.025). In Sec. IV, we turn to a closer investigation of the impact of the number of memories, and the details of the memory loss mechanism, on the dynamics.

B. Simulation results

Let us start with a simulation with $w = 40$, $\mu = 0.5$ and $N_h = 2 \times 10^4$. We assume that initially some number of individuals are infected by a single viral strain and there is a completely blank slate of memories for each of the hosts. Starting with 40 initial infecteds, the time evolution of the

TABLE I. Brief description of the model constituents.

Parameter/Symbol	Meaning (Default value)
x	Viral strain
$n(x)$	Number of infecteds with strain x
$I(x)$	Production rate of new infecteds with variant x
$V(x)$	Rate of emission of virion strain x
$E(x)$	Rate of virion emission by infecteds with strain x
$f_\epsilon(n(x))$	Cutoff function preventing infection when the number of infecteds when the local number of infecteds is too low
$p_{\rm inf}(j,x)$	The infection probability for host j exposed to strain x becoming infected
$g(\Delta x)$	The function governing the cross immunity between strains at a distance Δx
$mem(k,i)$	The ith memory of host k
$p_{inf}(x)$	The probability that a random host will become infected if exposed to strain x
$\rho_m(x)$	The population density of memories of strain x
$Q(x)$	An auxiliary field used to compute $p_{\rm inf}(x)$ from ρ_m
N_I	The total number of infecteds
N_h	The total number of hosts
R_0	The expected number of exposed hosts directly due to a single infection (3.8)
μ	Probability of an emitted virion being a different strain than its parent (0.025, 0.5)
M	Number of memories per host (6)
w	Width of cross immunity
ϵ	Cutoff local density below which infections do not occur
α	Exponent entering in the cut-off function f_ϵ (4)
p_0	A parameter quantifying the degree of partial immunity for the initial infecting strain (0.5)
χ	Fraction of noninfected hosts $\equiv 1 - N_I/N_h$
L_m	Length of mound
n_B	Density of infecteds in central region of mound
Q_B	Auxiliary field in central region of mound

system shows three distinct possibilities. One possibility is that the infected population collapses almost immediately; this happened for ten out of 50 independent simulations. This behavior occurs because viral phenotypic diffusion, as characterized by μ, does not occur fast enough to overcome immune suppression. The dynamics at short times resembles that of a pure single-variant Susceptible-Infected-Recovered (SIR) model, where the number of infecteds rises quickly and then falls to extinction, unless enough phenotypic diffusion occurs to escape the immune shadow of the initial variant. The percentage of immediate failures can be reduced by starting from a more spread-out initial state, but this is not the typical circumstance in realistic situations.

The other two behaviors are more interesting. One type is a relatively stable single propagating pulse [5,11]; this state was reached in approximately 50% of the simulations, but again this number depends on the detailed initial conditions. A snapshot of this behavior is shown in Fig. 2(a) along with the time history of the mean infection position along the phenotypic axis [Fig. 2(b)] and the history of $N_I(t)$ [Fig. 2(c)]. For this parameter set, the pulse, once formed, is very stable, lasting at least for many tens of thousands of simulation steps. This stability is due to the fact that the range of fluctuation as exhibited in Fig. 2(c) keeps the infected population number quite far from the absorbing state at $N_I = 0$. Thus, while the fluctuations are indeed relatively large, the qualitative behavior seen here is similar to what would be seen in a deterministic model exhibited, for example, in the work of Ref. [11]. Note that the average value of N_I is much smaller than N_h and so, for this state, model variants such as the cited ones that neglect the finite size host pool may be reasonable simplifications of our more complete formulation. The properties of this pulse behavior has been studied extensively [10,12] so we will not focus on it here.

Why does this pulse behavior arise in only 55% of the nonextinction runs? This arises from the fact that, in the absence of fluctuations, the single-variant initial condition gives rise to two counterpropagating pulses. If these two pulses grow to maturity, they eventually lead the way to a different dynamics, as we will explain directly below. Only if one and only one of the incipient pulses dies in infancy due to a fluctuation does a stable propagating pulse result.

So, what happens in the 40% of runs where two robust counterpropagating pulses are generated? The answer lies in an instability to the generation of new infections between the pulses. Since the two pulses consume twice as much of the memory as a single pulse, the degree of immune shadowing in the wake of the pulse is much reduced. Then, due to phenotypic diffusion, new infections generated behind the pulses are very likely to grow, in turn generating addition new pulses behind. This leads eventually to a broadly distributed infection pattern, as seen for example in Fig. 3(a). As more and more infecteds are generated between the original pulses, the number of susceptibles in the population is substantially reduced and the leading pulses at either edge slow down. The overall width of the pattern increases more and more slowly [Fig. 3(b)] until eventually saturating, even as the total infected population number remains relatively constant. As N_h is increased, an increasing percentage of runs result in this novel extended type of dynamical behavior. For example, for $N_h = 3 \times 10^4$, the number of runs leading to extinction is reduced to 4%, due to the reduction in the fluctuations leading

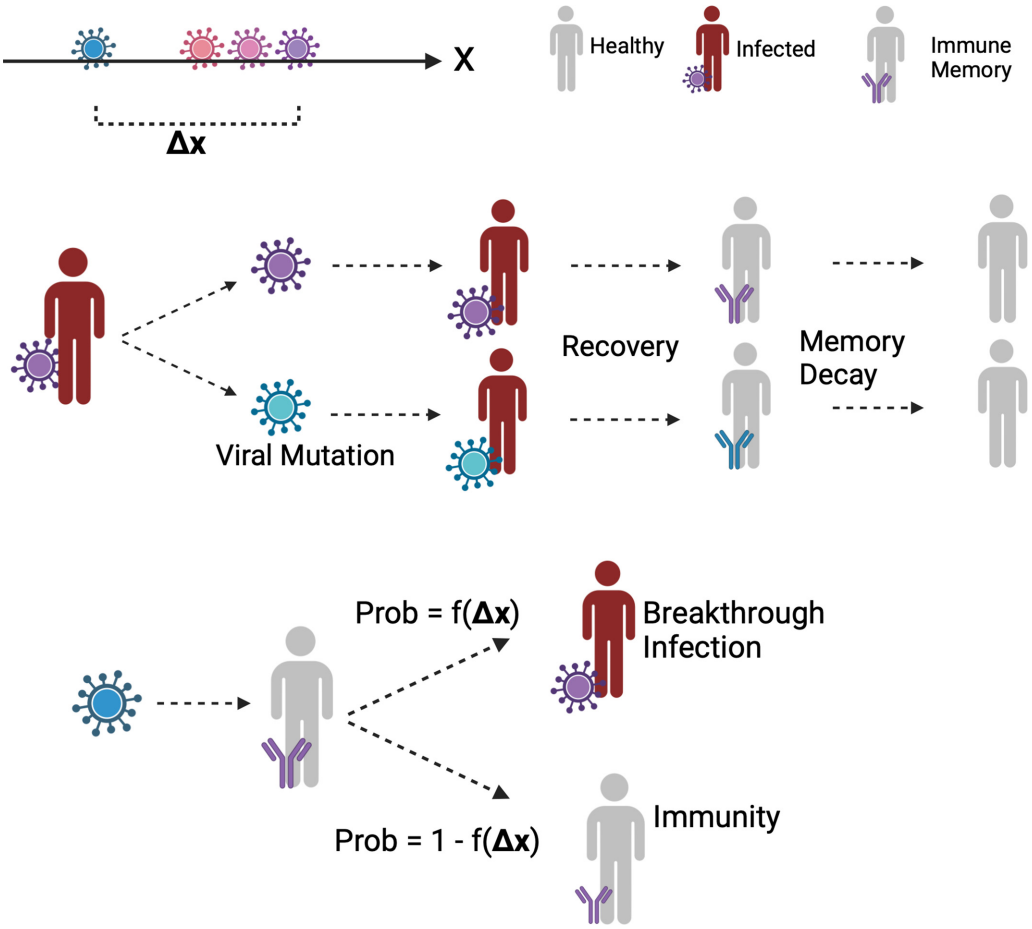

FIG. 1. A schematic view of the processes considered in our model. The space of allowed virions is taken to be one-dimensional, labeled by coordinate x. There is a release of virions by an infected person, some of which can be mutated by moving one unit in x. These virions then go on to infect other people, depending on the state of their immunity. Upon recovery, a previously infected person stores a memory of the infecting virus. Memories can be lost in a variety of ways, as discussed in the text. Finally, immune protection against one virus by the memory of another wanes as the distance between them increases.

to premature extinction. More dramatically, the number of runs leading to a single propagating pulse falls precipitously, to 24%, as fluctuation are what kill off one of the two pulses, leaving a single stable pulse. Thus, 72% of the runs produce a "mound" pattern.

What happens to the pulse solution if N_h is lowered? Carrying out simulations at $N_h = 12\,000$ reveals the fact that the pulse solution is in fact metastable. In Fig. 4(a) we plot $N_I(t)$ for a pulse that spontaneously collapsed after a significant amount of time. Clearly, going to a lower host population size has lowered the mean value of N_I for the pulse, thereby putting the absorbing state within the range of large but not excessively large fluctuations. The population depicted in the figure survived close calls at $t \sim 1900$, 2800, and 4000 before

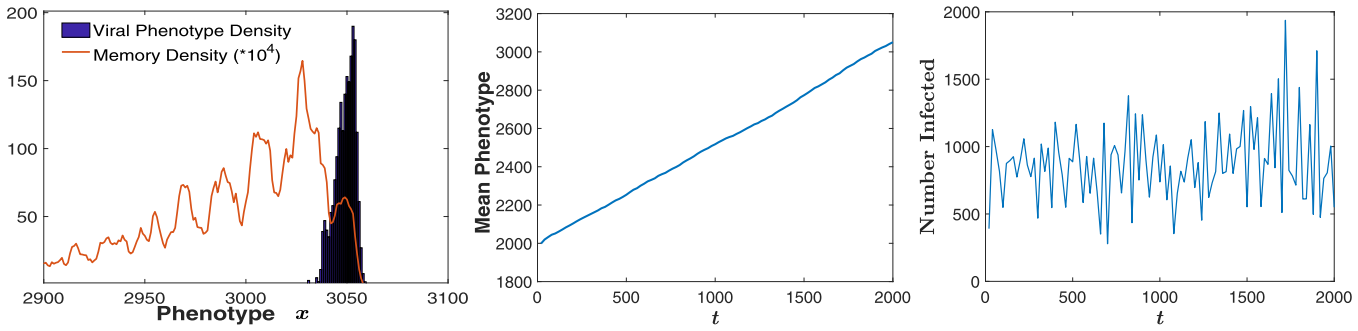

FIG. 2. A stochastic traveling pulse, for $w = 40$, $\mu = 0.5$, $N_h = 2 \times 10^4$. Left panel: An instantaneous snapshot of the viral strain density profile $n(x, t)$ and the memory density $\rho_m(x)$; Center panel: The mean of the positions (phenotypes) of the infected individuals, as a function of time t. Right panel: The total number of infected individuals, $N_I(t)$, as a function of time.

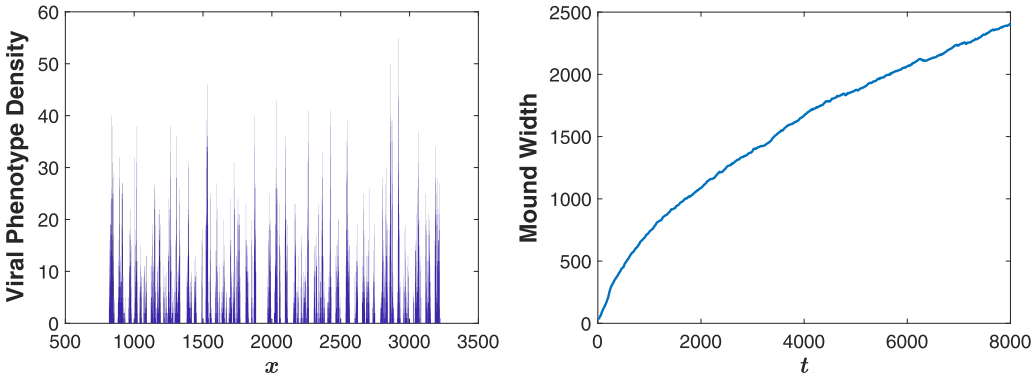

FIG. 3. The mound pattern for $w = 40$, $\mu = 0.5$ and $N_h = 2 \times 10^4$. Left panel: An instantaneous snapshot of the viral phenotype densities $n(x, t)$, showing an extended tail in the wake of an originally propagating pulse. Right panel: The mound width, namely the distance between the largest and small positions of individuals in the mound, as a function of t.

finally succumbing at $t \sim 4400$. Our data is consistent with a constant rate of decay of the probability of survival of the pulse, which of course implies an exponential distribution of survival times. The decay rate appears to be a strongly decreasing function of N_h accounting for the lack of observations of the long-time pulse collapse in our simulations at $N_h =$ 20 K. The main conclusion to be drawn from these simulations is that while long-lived stable pulses are a possible outcome with a localized initial infection, they are by no means the only outcome, together with extinction and mounds, with extinction dominant for smaller N_h and mounds for larger N_h. Hence, investigating the nature of the mound pattern in more detail is a worthwhile task, to which we shortly turn.

Before proceeding, though, let us make a few observations about multipulse solutions. Given that at sufficiently large N, a single pulse solution exists and is linearly stable, one might expect multipulse solutions with either counter-propagating or copropagating infected host populations which are far enough apart to minimize cross reactivity. Figure 3(b) shows an example of a counterpropagating pair, at the much lower value of $N_h = 5000$. Interestingly, this pattern is not typically selected from generic localized initial conditions, unlike what is observed in excitable media with purely local inhibition where each edge of the initial data generates an independent and oppositely propagating pulse. Furthermore, no matter how far apart the pulses are, they continue to interact via the limitations on immune memory; a host becoming infected by a far-away virion will degrade its immune response to the near virion by which it was most recently infected. In other words, the model has long-range coupling in phenotypic space. This tends to diminish the ability of the pulse wake to remain stable and thereby lowers the range of the overall host population size for which the two-pulse solution is stable.

We also carried out calculations at values of the parameters which may be more physiological and which more strongly emphasize stochasticity. Specifically, we set $w = 10$ which makes the immune shadow of an infection pattern much less spatially extended. To partially compensate, we set $\mu =$ 0.025; note that this is close to the naive estimate that one can change these parameters while maintaining the dimensionless ratio $\tilde{\mu} = \mu/w^2$ constant and expect only a limited change of behavior. The key finding is that the basic phenomenology is unchanged, and so the basic range of behaviors presented is independent of the details of the parameter choice. Thus, for $N_h = 2 \times 10^4$, we get mound patterns in 74% of the runs and pulses in the rest. For $N = 1.5 \times 10^4$, we get mounds in 52% of the runs and pulses in the rest, one of which dies at $t = 650$ due to a fluctuation. Reducing N_h further to $N_h = 10^4$, we encounter 12% immediate (before $t = 100$) extinctions, 22% mound patterns, and the rest pulses, 1/3 of which die

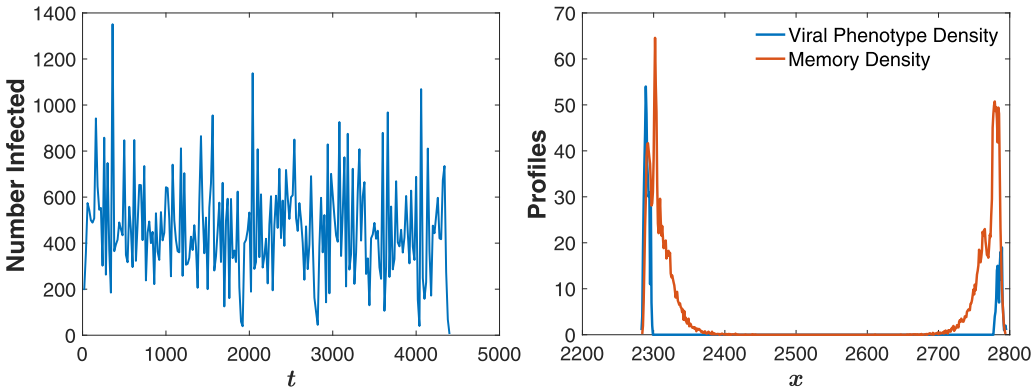

FIG. 4. Left panel: The total number of infected hosts, $N_I(t)$, showing the collapse of a traveling pulse for $N_h = 1.2 \times 10^4$. Right panel: Two counterpropagating pulses for $N_h = 5000$, showing the viral phenotype profile $n(x)$ in black and the (unnormalized) memory profile $\rho_M(k) \equiv \sum_{i,j} \delta_{mem(i,j),k}$ in red. For both panels, $w = 40$ and $\mu = 0.5$.

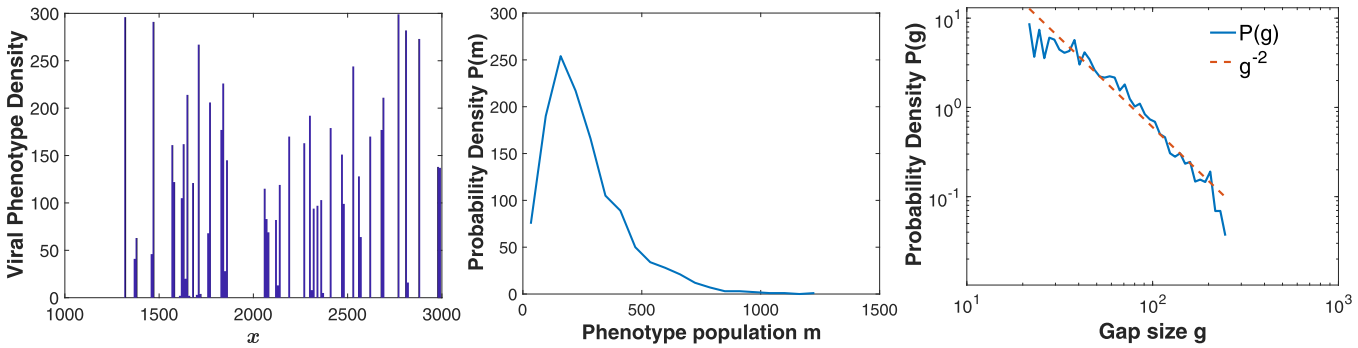

FIG. 5. Stochastic simulation for $w = 10$, $\mu = 0.025$, $N_h = 10^4$. Left panel: Infection profile $n(x)$. Center panel: Distribution of patch (defined as a set of strains without a gap of $2w$ separating them, so that it constitutes an effective phenotype) population sizes $P(m)$. Right panel: Distribution of gap sizes, $P(g)$, between effective phenotypes, showing an approximate $P(g) \sim g^{-2}$ power law.

due to a fluctuation before $t = 1000$. Thus, the general trend that decreasing N_h gives rise both to more immediate extinctions as well as fully developed pulses, whose average lifetime decreases with decreasing N_h. In the next subsection, we will discuss the mound pattern in more detail, examining its development in time and its internal structure.

C. More on mounds

As already mentioned, simulations at large N_h typically lead to extended patterns, with localized patches of infection separated by gaps, i.e., ranges of phenotypes not present in the population. As has already been seen in Fig. 3(b), the width of the overall structure increases slowly over time and takes a long time to reach a very noisy steady state. Since the overall population remains roughly constant (at approximately $2/3N_h$), this implies that either the patch size must be decreasing or the typical gap must be widening during this long transient. Now, just as we saw for isolated pulses, patches cannot remain stable for long times when their size becomes too small. Hence, as patches shrink, they will disappear and lead to larger gaps between neighboring patches. This is balanced initially by the creation of new patches in the wake of existing ones (see below), but this effect diminishes as the number of patches becomes large and N_I reaches its steady-state value much above that of a single pulse. Also, patches have a small repulsive interaction due to residual immune memory effects. All this means that over the transient period, the major effect is a spreading apart of the patches. This will only cease when the rate at which the outermost patches disappear matches the average rate at which they move outward. This behavior leads to a very diffuse pattern in the very long time limit.

In Fig. 5(a), we present a second example of this structure, at $N_h = 10^4$, $w = 10$ and $\mu = 0.025$, now focusing on simulation times long enough to have reached the very noisy steady state. As advertised, the basic structure is that of relatively isolated patches of infected host density separated by gaps that are much larger than w. In Figs. 5(b) and 5(c) we plot, respectively, the distribution of patch populations and of the gaps. The patch population size exhibits a peaked distribution and hence the patches can be characterized by a typical population size. On the other hand, the gap distribution is a power law, corresponding to an extremely disordered pattern. The exponent in the power varies as a function of parameters, i.e., is not universal (data not shown).

As the pulse pattern and the extended endemic pattern can both occur at the same set of parameters, the emergence of the latter cannot be directly due to a linear instability of the former. Instead, there appears to be a finite amplitude instability wherein a nascent peak appears in the wake of a propagating pulse and manages to overcome the immune system inhibition. Once this occurs for the leading peak, the system cascades into the full extended pattern; this presumably occurs because the effective immune "shadow" decreases with the number of patches as they compete for immune memory slots, as explained above for a two-pulse solution. This behavior gives rise to a rapid expansion phase, which only reaches the slow transient behavior described above when N_I has reached its asymptotic value and the creation of new pulses becomes much less common.

It is of course quite difficult to devise analytic approaches to a complex multifield nonlinear spatially extended stochastic dynamical system. To make further progress, we next introduce a deterministic approximation to the dynamics, as already has been proven useful for the stable pulse solution in related models [5,6,11]. Our emphasis therefore will be on the extended pattern, whose existence and properties constitute the major new results of this work.

III. DETERMINISTIC EQUATIONS

A. Derivation

We now proceed to derive a mean-field deterministic model that will allow us to gain some analytic insight into the coevolutionary dynamics. We will work in the spatial continuum limit where differences and sums are replaced by derivatives and integrals. The equation for the density of infected hosts is just

$$\dot{n}(x) = -rn(x) + I(x), \tag{2}$$

where r is the recovery rate and I is the rate of new infections of type x. The I field in turn is given as the product of three factors,

$$I(x) = V(x)p_{\rm inf}(x)\left(1 - \frac{N_I}{N_h}\right). \tag{3}$$

Here, $V(x)$ is the rate of virion emission including the effects of mutation, while $p_{\rm inf}(x)$ determines the immune inhibition. The last factor accounts for the fact that currently infected people cannot acquire a new infection. In our stochastic model, the mean rate of virion emission by infected hosts of type x is given by $rR_0n(x)$. In constructing a deterministic model, we have also to incorporate the well-established idea [4,5,19–21] that the production of new virions needs to be cut off at small n as an approximate way to incorporate the discreteness of individual hosts. Thus the rate of emission by infected hosts carrying strain x is taken to be

$$E(x) = rR_0n(x)f_\epsilon(n(x)), \tag{4}$$

where f_ϵ is a function that goes from 0 to 1 as its argument increases and whose natural width depends on the cutoff parameter ϵ. In particular, we choose the cutoff function

$$f_\epsilon(n) = 1 - e^{-(\frac{n}{\epsilon})^\alpha}. \tag{5}$$

This function vanishes as n^α for $n \ll \epsilon$ and approaches 1 exponentially for $n > \epsilon$.

In our stochastic model, newly created virions can, with probability μ, represent a strain mutated with respect to their parent virus strain x, with their new phenotype being with equal probability either $x-1$ or $x+1$. This translates to a total viral emission rate for strain x of

$$V(x) = E(x) + \frac{\mu}{2}E''. \tag{6}$$

Note that in this continuum formulation, the only two quantities with dimensions of length are $\sqrt{\mu}$ and w, which were in the stochastic, lattice formulation, measured in units where the mutational change in x was unity. Thus, the relevant dimension-free parameter is the ratio μ/w^2.

The remaining challenge is to devise a deterministic treatment of the immune memory and the concomitant construction of $p_{\rm inf}$. We assume that there is no correlation between memories in the different slots of a given host or between hosts, so that the state of the memory is characterized simply by the memory density function $\rho_M(x,t)$. With this assumption,

$$p_{\rm inf}(x) = \left(\int dx'\ \rho_M(x')(1 - p_0 g(|x-x'|))\right)^M, \tag{7}$$

where, as before, g is an exponential with width w. Our stochastic simulations show that this is a quite good assumption, at least for $M = 6$. Finally, we need to update the memory density. This occurs whenever a host recovers, and the resultant equation is

$$\begin{aligned}\dot\rho_M(x) &= -\frac{\rho_M(x)}{MN_h}\int dx'\ rn(x') + \frac{rn(x)}{MN_h}\\ &= -\frac{r\rho_M(x)N_I}{MN_h} + \frac{rn(x)}{MN_h}.\end{aligned} \tag{8}$$

The first term accounts for the overwriting of old memories, as each memory accounts for a fraction $1/(MN_h)$ in the density. The second term refers to the writing of a new memory; note that the memory density remains normalized $\int \rho_M(x)dx = 1$, as the time derivative of this sum automatically vanishes.

The fact that the infection probability is a nonlocal function of the memory density is an annoyance when it comes to both numerical calculation and analysis. It is, therefore, more convenient to define an auxiliary field Q via

$$Q'' - \frac{Q}{w^2} = -\frac{2}{w}\rho_M, \tag{9}$$

supplemented by the boundary conditions that ensure that

$$Q(x) = \left(\int dx'\rho_M(x')g(|x-x'|)\right), \tag{10}$$

in terms of which

$$p_{\rm inf}(x) = (1 - p_0Q(x))^M. \tag{11}$$

Q has no independent physical significance, as indicated by the fact that it has no actual dynamics; it is merely a way to track the factor by which infectivity is reduced at position x by the integrated effect of all the current memories.

We thus obtain the two dynamical equations in their final form

$$\dot n = -rn + I = -rn + V(1 - p_0Q)^M, \tag{12a}$$

$$\dot\rho_M = -\frac{r}{N_hM}(\rho_MN_I - n) \tag{12b}$$

with the auxiliary equations

$$E = rR_0n\left(1 - \frac{N_I}{N_h}\right)f_\epsilon(n), \tag{13a}$$

$$V = E + \frac{\mu}{2}E'', \tag{13b}$$

$$Q'' - \frac{Q}{w^2} = -\frac{2}{w}\rho_M. \tag{13c}$$

Finally, an explicit global coupling in the model is through the dependence on N_I, the total number of infected individuals at a specific time, which enters into the above determination of the emitted virion field E. Of course, there is also global coupling through the Q field.

B. Pulse solution

As a warm-up exercise, we first proceed to understand why the deterministic equation allows for an isolated pulse solution, moving with a steady-state shape at a specific constant velocity c and containing a precise number of infected hosts. To see this, we need to analyze the deterministic equation in the two asymptotic regimes; the tail behind the front and the leading edge ahead of it. We can thereafter employ the idea of counting the number of parameters that need to be adjusted to satisfy matching conditions in order to determine the dimension of the space of allowed solutions. As the derivation is fairly involved, we sketch it here and relegate the details to Appendix A.

The lack of new infections in the tail implies that the infected host density decays exponentially to the left (assuming a rightward propagating pulse), with an undetermined magnitude. From Eq. (13a), we get that E likewise decays exponentially, with a magnitude set by that of n. From E we can uniquely determine V from Eq. (13b). Also ρ_m as given by Eq. (12b) has a homogeneous exponentially decaying solution, with an arbitrary amplitude as well as an inhomogeneous term driven by n. Lastly, Q has a homogeneous term, with undetermined amplitude, that can be added to the particular

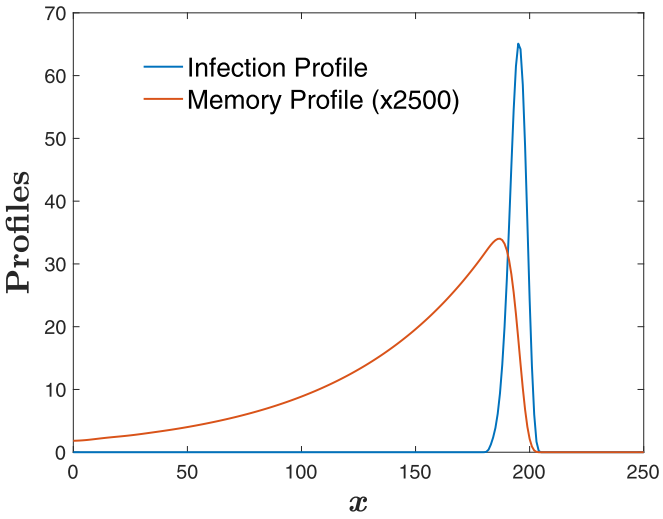

FIG. 6. Infection profile $n(x)$ and memory density (rescaled by a factor of 2500) $\rho_M(x)$ for a deterministic pulse, with $w = 40$, $\mu = 0.5$, $N_h = 15000$, $\epsilon = 1$, $\alpha = 4$.

solution driven by ρ_m. Thus the most general solution in the tail has three unknown constants, the amplitudes of n and the homogeneous pieces of ρ_m and Q.

The situation is considerably more complicated at the leading edge, as here we have new infections occurring. The cutoff in Eq. (12a) plays a key role here, as mentioned above. Since the cutoff not only multiplies the E term in V, but also the "diffusion" term E'', we effectively have a case of nonlinear diffusion. Hence we might expect (and will verify below) that solutions exist for which the infected host population becomes precisely zero after some value of x. This is analogous to what has been seen in the Fisher equation with nonlinear diffusivity [22–24]; in particular, it has been shown there that the fronts that form from localized initial conditions have this specific property. Thus, we have to first find the specific form of this nonlinear solution and then determine, by linearizing away from that solution, the number of unknowns governing the fields in this region. This calculation is presented in Appendix A, with the result that there is one unknown in this region, specifically Q evaluated at the leading edge.

So, our two asymptotic solutions have a total of four unknowns, and adding in the velocity c makes five. This needs to be compared to the number of matching conditions if the equations are integrated separately from the two edges to some common point in the middle. Q satisfies a second-order equation and hence both Q and Q' must be continuous. ρ_M satisfies a first-order equation and hence it needs to be continuous. Since I depends on two derivatives of n, the n equation is effectively second-order; hence we must impose that n and n' are continuous. This gives a total of five equations, showing that in general there will be a discrete set of pulse solutions. Numerically, there appears to be only one stable solution of this type. A simulation starting from a localized source will indeed converge to such a solution. An example of a pulse solution is shown in Fig. 6, where both the infected host population n and the memory density ρ_M are presented. We can explicitly check that this solution quantitatively agrees with the exponential behavior in the tail. The situation at the leading edge is trickier because of the singular behavior there. One must be careful to avoid effects due to the discrete nature of the phenotypic space (and of the simulation algorithm) in order to see the structure predicted by our spatially continuous treatment. A simple approach yields qualitative but not extremely quantitative agreement. We will carry out this comparison in the next section, for a different type of solution.

As already mentioned, the single pulse solution in this model is similar in many respects to pulses found in other evolutionary dynamics problems. There are of course residual questions that are worth investigating, especially the dependence of the total infected number N_I on the host population size N_h. However, the major point of this work is to understand the more exotic dynamical behavior found in the stochastic simulations, namely the extended patterns that emerge due to the ineffectiveness of the immune system in suppressing fluctuations that arise in the wake of a propagating pulse. This happens for large N_h, which translated in the deterministic model should correspond to small ϵ. We now turn to a discussion of this behavior.

C. Mounds

In Fig. 7, we present a series of snapshots from a simulation of the deterministic model at a low cutoff value. Initially a pulse is formed, but relatively quickly a second peak of the infected population density appears in the wake of the original pulse. This process can be seen in Fig. 7(a), where the second peak has appeared due to the failure of the immune "shadow" to adequately suppress its growth. This instability leads to a new type of static solution which we refer to as the mound. For these values of the parameters, this is a finite amplitude instability, as a careful preparation of the pulse state allows it to propagate indefinitely without this type of instability.

After the initial instability, the system settles into a slowly expanding pattern [Figs. 7(b) and 7(c)] with two counterpropagating peaks on the outside leaving behind a constant infected host density region between them, the "bulk" region, which we denote by n_B. Similarly, in this region, the memory density is also essentially constant, $\rho_M \approx \rho_{M_B}$, and so is the associated auxilliary field $Q \approx Q_B$. Let us assume that we can ignore for the moment the slow spreading of this constant region. Then, the bulk state obeys the time- and space-independent equations

$$\begin{aligned} rn_B &= E(1 - p_0 Q_B)^M \chi, \\ \frac{Q_B}{w^2} &= \frac{2}{w}\rho_{M,B}, \\ \rho_{M,B} &= \frac{I(x)}{I_T} = \frac{n_B}{N_I}, \\ E &= rR_0 n_B f_\epsilon(n_B), \end{aligned} \tag{14}$$

where for convenience we have introduced $\chi \equiv 1 - N_I/N_h$, the fraction of uninfected hosts. The solution of this set of equations determines the bulk infected density n_B via the implicit condition

$$\left(1 - 2w\frac{n_B}{N_I}p_0\right)^M = \frac{1}{R_0\chi}. \tag{15}$$

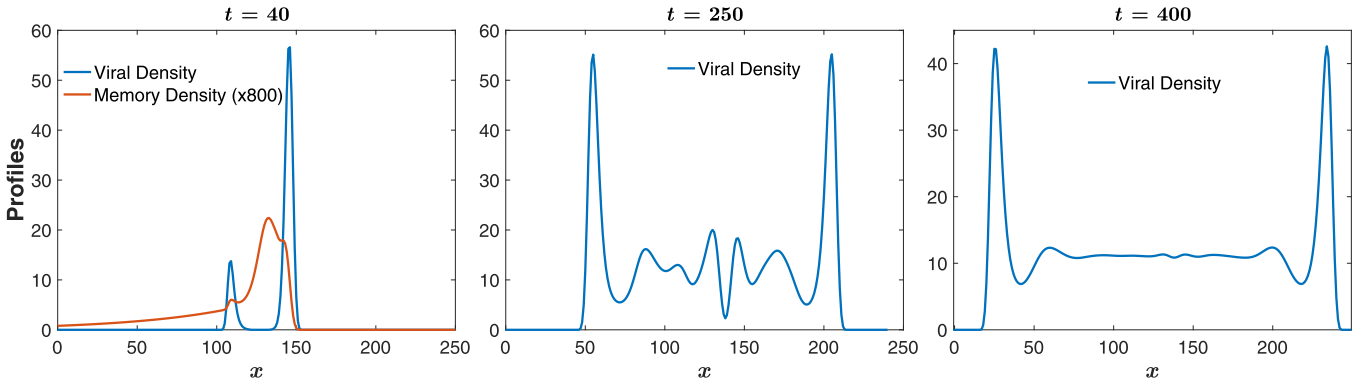

FIG. 7. Time development of a deterministic mound solution, for $w = 20$, $\mu = 0.125$, $N_h = 5000$, $\epsilon = 0.05$, $\alpha = 4$. Left panel: Infection profile $n(x)$ (blue) and rescaled Memory profile $\rho_M(x)$ (red), at time $t = 40$. Center panel: Infection profile $n(x)$ for $t = 250$. Right panel: Infection profile $n(x)$ for $t = 400$.

This is valid in the case where n_B is large enough such that the cutoff function can be taken to equal unity. Now, while the solution is expanding, we clearly have $N_I \approx n_B L_m$, where $L - m$ is the instantaneous length of the mound (or, more precisely, of the bulk region), so that $\chi \approx 1 - n_B L_m / N_h$. Thus, the above equation gives rise to the simple predictions,

$$n_B(t) = \frac{N_h}{L_m R_0}\left[R_0 - \left(1 - \frac{2w}{L_m(t)}p_0\right)^{-M}\right]$$
$$Q_B(t) = \frac{2w}{L_m(t)}, \tag{16}$$

so that n_B slowly decreases as the mound expands. Once we have this result, we can qualitatively understand the spreading dynamics. The pair of pulses delineating the bulk regime will each move at a velocity v that can be found from the type of two-pulse solution we have seen in the stochastic model, Fig. 4(b); with $\frac{dL_m}{dt} = 2v$. Here, however, the parameter χ, which is related to the total number of infected hosts, and the global field Q are no longer determined by the pulse solution itself, but instead are dominated by the bulk region. The reduction of χ as N_I increases will slow the pulse, so that the velocity is time varying. A comparison of these predictions with our numerical results is presented in the first two panels of Fig. 8. In the leftmost panel, our above prediction for n_B based on the measured $L_m(t)$ is compared to the measured value of $n(t)$ at the center of the mound. In the middle panel, our prediction for Q_B based on the measured $L_m(t)$ is tested against the direct measurement of $Q(x)$ at the mound center. Both these panels show good agreement at large t. The rightmost panel shows the falloff of the velocity at large times.

If there were no cutoff, the above argument suggests that the pulse would spread forever and would never reach a time-independent solution. This is of course not what we found for the stochastic model, where the system does eventually reach a statistical steady state. The cause of this stopping is that eventually the cutoff becomes important in letting the pulse velocity go to zero. To understand this in more detail, we can consider the slightly simplified problem of an infinitely-long static "half-mound" which connects the aforementioned static solution as $x \to -\infty$ with a nonpropagating pulse which completely vanishes at some particular x value (again chosen to be $x = 1$), as did the propagating pulse solution studied above. In this solution, the parameter χ is taken to be a free parameter and it directly determines N_I through $N_I = N_h(1 - \chi)$. Once the solution is found, however, it can be reinterpreted as a large but finite solution where the length L_m is determined by the consistency requirement which is approximately $N_I = n_0 L_m$ or more exactly that the assumed value of N_I is actual equal to $2\int_{1-L_m/2}^{1} n(x)dx$.

As with the pulse solution, we need to analyze separately the two asymptotic regimes, the semi-infinite bulk and the

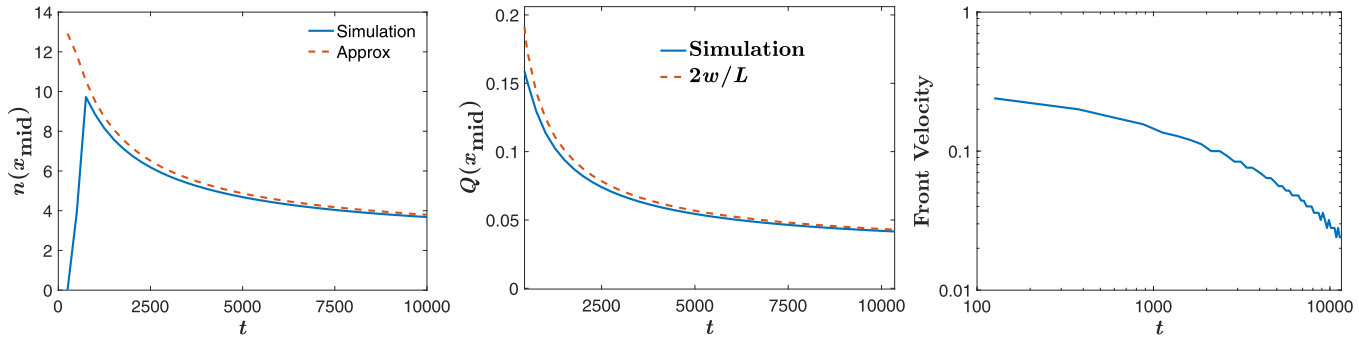

FIG. 8. Comparison of numerical simulation with asymptotic theory from Eq. (16). The left block shows $n(x_{\rm mid}$, the infected population at xmid, the middle of the mound, the middle shows $Q(x_{\rm mid}$,the auxiliary field Q at $x_{\rm mid}$, and the right is a plot of the decreasing velocity, showing the decay of the mound front velocity toward zero.

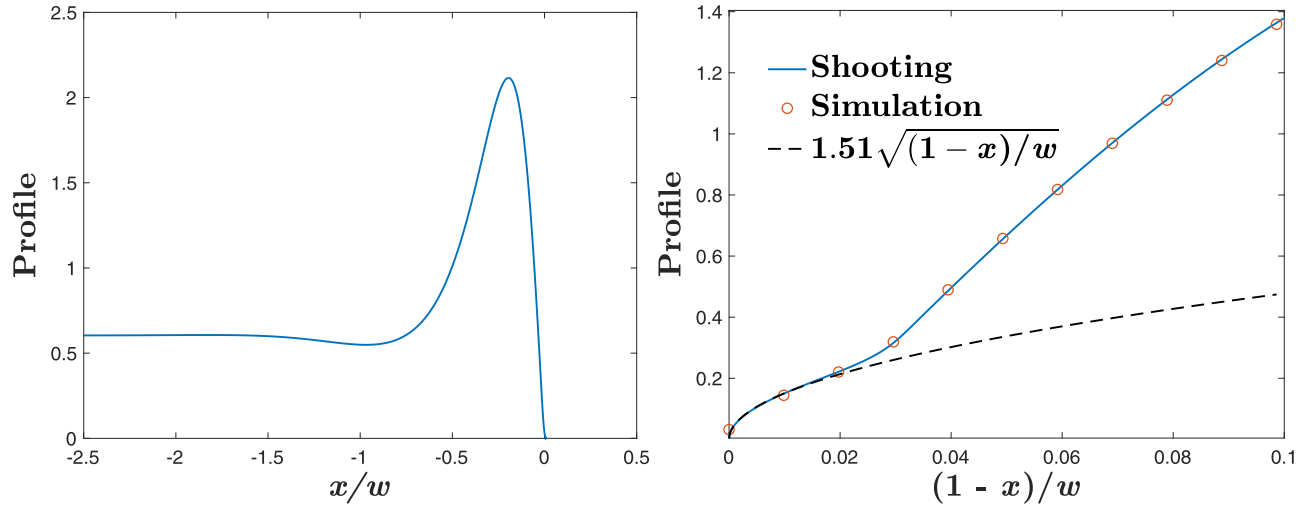

FIG. 9. Left panel: The half-infinite mound infection profile $n(x)$ as found by shooting for $\mu/w^2 = 2.5 \times 10^{-4}$, $N_h = 5000$, $\epsilon = 0.22$, $\alpha = 4$. Right panel: A blow-up of the leading edge of the half-mound profile, along with the tip region of the stationary mound obtained from a dynamical simulation on a discrete lattice and also the predicted $\sqrt{1-x}$ behavior.

front, determining the degrees of freedom in both regions. We first consider the bulk region, considering spatially dependent deviations from the constant bulk solution. Restoring the spatial derivatives and expanding around the solution, we have for the deviations of the various fields from their asymptotic limits,

$$\delta_n = \delta_n\left(1 + \frac{\mu}{2}k^2\right) - p_0 n_0 M (1 - p_0 Q_0)^{-1}\delta_Q,$$
$$\delta_Q\left(k^2 - \frac{1}{w^2}\right) = -\frac{2}{wN_I}\delta_n. \tag{17}$$

This yields the dispersion relation

$$(w^2k^2 - 1)k^2 = -\frac{p_0 n_0 M w}{\mu N_I (1 - p_0 Q_0)}. \tag{18}$$

In general, this gives rise to four complex modes of the form $e^{\pm ax \pm ibx}$. Thus, as $x \to -\infty$, there are two bad and two good modes, implying that the solution for large negative x has two unknown parameters, namely the coefficients of the two good modes.

We must next turn to the leading edge. This is presented in detail in Appendix B. The result is the same as before, namely that the only new degree of freedom is the value of Q. We then are finally in a position to count degrees of freedom and compare to the continuity conditions that must be imposed to connect the two asymptotic regimes, say at $x = 0$. As compared to the propagating pulse, there is one fewer matching condition, namely four here, as the order of the n equation has been reduced by 1. The number of unknowns is similarly four; two coefficients at large negative x, $Q(1)$ and χ. Thus there is expected to be a unique solution, or perhaps a discrete set of solutions. Our simulations strongly suggest that there is only one such solution that is dynamically stable, as the system is always attracted to a unique mound state.

D. Shooting method

We have implemented a shooting method to solve for the half-infinite mound. Based on the above analysis, we shoot forward from an arbitrarily chosen x_b, where we choose a small amplitude perturbation of the bulk solution. The perturbation is characterized by two matching parameters, the phase of the growing fluctuation and N_I. The fourth-order set of ordinary differential equations is integrated forward to the point where $n'(x)$ vanishes and n is sufficiently large that we are out of the tail region. In parallel, we shoot backward from the point $x = 1$ where the density vanishes. Here, the matching parameters are $Q(1)$ and N_I. Again, we integrate until $n'(x)$ first vanishes. At this point, the other three fields have to agree, giving us three equations in three unknowns.

The resulting numerical solution is plotted in Fig. 9 and matches exactly that produced by the time-dependent simulation. In particular, the behavior of the simulated system at the $x = 1$ singular point exactly matches that predicted by the analytical treatment. To make this occur, we had to choose a very large value of μ so as to limit the effects of phenotypic discreteness (i.e., the spatial lattice with constant one) introduced by the finite difference simulation strategy. At smaller values of μ, the numerical solutions become affected by the lattice and the quantitative (but not qualitative) level of agreement degrades.

E. Combs

There is one additional pattern that is possible in the deterministic system. Figure 10 shows what we will call a "comb" configuration. This is a time-independent solution of the governing equations that again emerges via an instability of the tail of an originally propagating pulse. Here, however, the immune shadow behind each peak is large enough to prevent the full mound from forming, thereby leaving behind a pattern of isolated peaks.

Comparing the deterministic findings to the stochastic simulations, we see rough congruence. Of course, the deterministic system can never exhibit a complete collapse of the infected population; this collapse is a fluctuation-induced effect that ultimately depends on the absorbing state nature of the $n = 0$ solution. Aside from this, we saw that both models

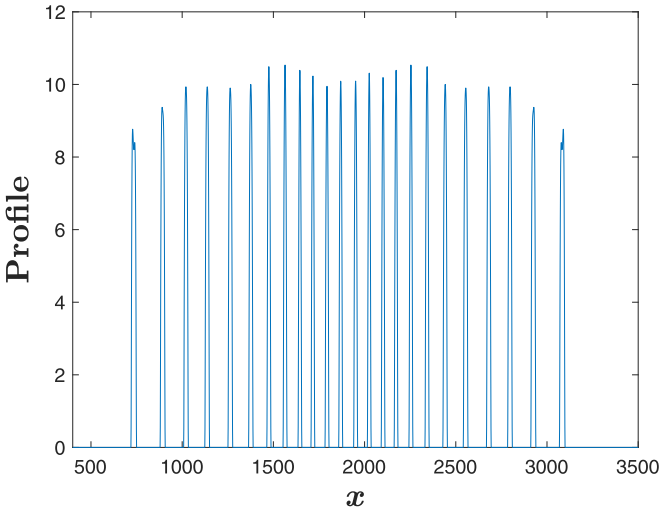

FIG. 10. The deterministic comb pattern infection profile $n(x)$ for $w = 40$, $\mu = 0.5$, $N_h = 5500$, $\epsilon = 1$, $\alpha = 4$.

exhibit two types of states, a propagating local pulse versus an extended, eventually statistically steady, configuration. The difference between the deterministic mound and comb states has been obfuscated in the stochastic dynamics. The time average of the stochastic simulation is clearly consistent with the mound state even though a single snapshot is perhaps more analogous to a comb structure with its isolated peaks. For both cases, the number of infected hosts is a significant percentage of the total number of those present, unlike the case of the pulse solution with N_I much less than N_h.

IV. THE ROLE OF MEMORY DECAY

In the previous sections, we have demonstrated the existence of a tail instability that often leads to the formation of an endemic state of infection. This was shown initially in the full stochastic model and then analyzed in detail using a deterministic PDE approximation thereof. We note in passing that this instability was eliminated by hand in Ref. [10] by choosing a highly nonsymmetric infectivity profile which for a pulse never forgets previous infections. The physical justification for such a kernel is not clear and in any case does not appear to accurately describe respiratory infections (see below). A similar *ad hoc* modification of the model was responsible for eliminating the instability in Ref. [12]. This instability arises from the decay of memories in our model due to the finite size of the memory buffer. We have checked that there can be no such instability if we allow an unbounded set of memories. We also checked that increasing the memory size will push the instability to larger values of N_h, assuming that all other parameters remain unchanged. In other words, there needs to be more available hosts to support the endemic state if each host has a more effective immune memory system.

The assumption that the new memory induced by the recovery from a new infection must replace an existing one is perhaps most closely realized in the CRISPR-based bacterial immune context [25]. There, virus information is stored in a DNA array which remains roughly at fixed size even as new viruses need to be remembered [26]. There is still some uncertainty as to the precise molecular mechanism that enables this behavior and whether there is any bias in which memory is deleted to allow room for a new one to be stored. We have therefore checked that one finds an analogous tail instability (leading again to an endemic state) if one changes our model so as to drop the oldest rather than a randomly chosen memory. An example of the still-evolving endemic state along with the relevant infectivity profile is presented in Fig. 10. Going to this new model variant lowers the memory size threshold above which the tail instability vanishes. This occurs because dropping the oldest memory makes the immune system more potent to the variants currently dominant in the system.

What about the human adaptive immune system and other similar ones? There is significant evidence that at least for some respiratory viruses such as the SARS family, there is indeed a decay of immune response over extended times [27,28]. This decay apparently involves the reduction of the number of long-lived antibodies, as has been shown for SARS directly and more recently for COVID-19 [29]. It does not appear to be true for other viral diseases such as measles. Thus, diseases with memory loss typically lead to the SEIRS class of models [30], corresponding to susceptible, exposed, infected, recovered, and returning to susceptible subpopulations. Our model is of course a variant-resolved version of SEIRS, albeit with a more complex return to susceptibility depending (in our case) on the degree of mutations.

To test whether the precise form used above, involving memory replacement as opposed to simple loss, is essential, we can instead introduce memory decay as a simple stochastic process. Here, we eliminate memory replacement and in its place introduce a rate at which an individual memory decays. This additional stochastic process becomes part of the overall Gillespie algorithm, which now chooses appropriately between individual host recovery, new infection of a host, and decay of a randomly chosen memory. In Fig. 11, we present simulation snapshots from this new variant model, first at the ratio of memory decay to host recovery of 0.005 (showing no tail instability) and then at 0.03 (showing a tail instability). These results clearly indicate that the precise form of memory decay is not an essential aspect of the new endemic state. We can therefore expect that both respiratory viruses in mammals as well as phages in bacteria might exhibit this possibility.

For this set of parameters, a ratio of 0.01 is at the margin, which means that some runs result in extended states and some do not. Interestingly, this value seems to be in the relevant physiological range for COVID-19 [29], several years versus a week or two. Basically, the issue is the extent to which individuals left behind the main pulse remain infected long enough for their viral loads to take advantage of memory decay in the population and ignite a new burst. This new burst must rely on increased infectivity and if we compare the two subfigures of Fig. 12, we see the difference in the infectivity (roughly 0.3 versus 0.1) in the unstable versus stable cases. We also point out that we study the case of two counterpropagating pulses due to convenience, as this is the state that is easiest to produce starting from a localized infected host population as an initial condition. All of our findings equally apply to one propagating pulse, albeit with some change (i.e., lowering) in

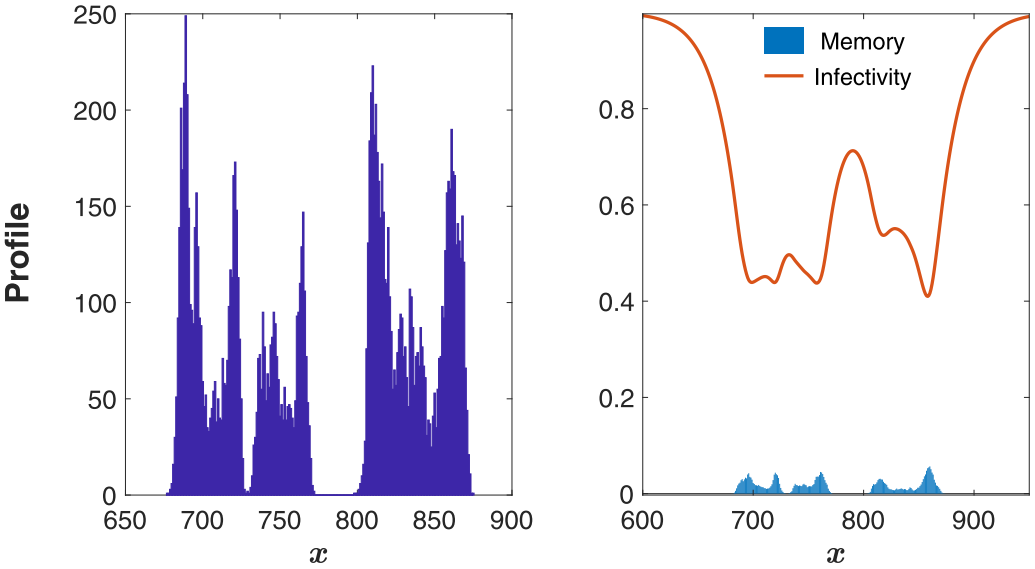

FIG. 11. Simulation of a case where the oldest memory is replaced upon new infection. Here, a snapshot of the nascent mound state is shown, for $N_h = 30$ K, $w = 20$, $\mu = 0.5$ and a memory buffer of size 3.

the stability threshold due to the elimination of the requirement for hosts to remember two separate coexisting strains.

V. DISCUSSION

Understanding the never-ending battle between viruses and our adaptive immune system is clearly a critical issue in modern-day statistical physics and biomedicine. The problem and its importance have been amply demonstrated by our experience with COVID-19. A new virus emerges, and the immune response is constantly challenged by the ability of the virus to mutate to new variant strains that are beyond the range of immune coverage. We certainly need to do a better job of understanding the resultant dynamics and devising possible ways to mitigate the corresponding consequences.

Viral evolution has been effectively modeled by devising equations that govern population propagation in a fitness landscape. These ideas have been successfully applied both to artificial evolution in laboratory experiments as well as to HIV infections. However, most of these works have assumed that the aforementioned fitness landscape, however complex, was not able to adapt to changing viral populations. This is explicitly not the case when considering the role of the immune system in responding to viral infection. Thus, several recent papers on coupling viral evolution to dynamic immune system response represent a welcome expansion of the applicability of fitness evolution modeling.

This work has presented a new formulation of viral-immune coevolution, including the effects of a finite host population and the dynamics of a (possibly) finite immune

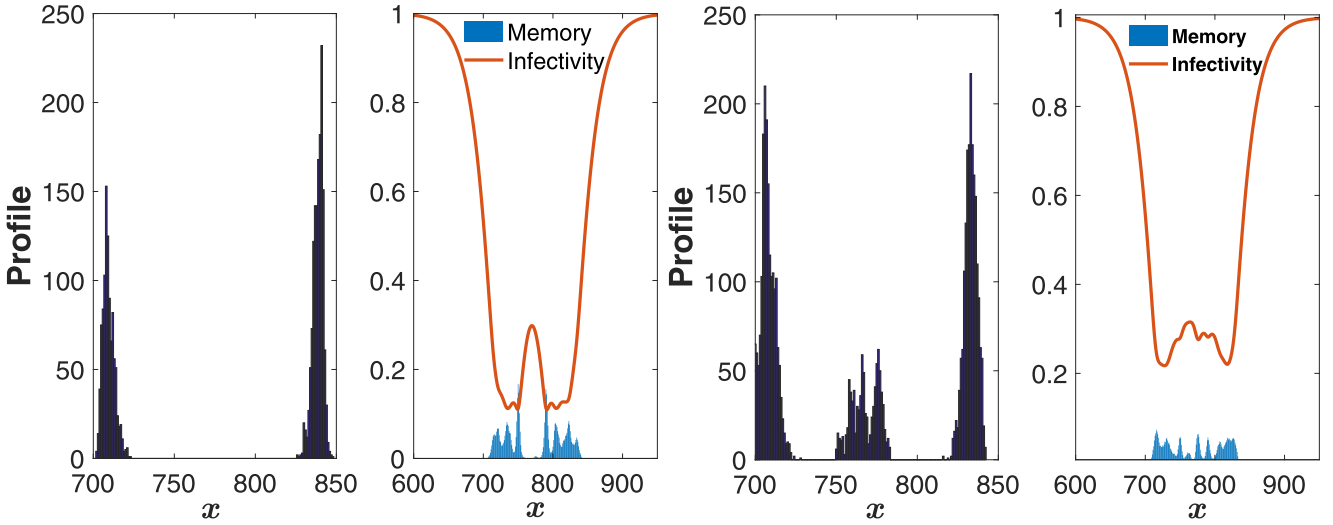

FIG. 12. Simulations of the memory decay variant where there is no *a priori* limit on memory size, but individual memories decay with variable rates. Other parameters are the same as in Fig. 11. In the left two panels, the ratio of memory decay rate to infection recovery rate is 0.005, and in the right two panels it is 0.03. Note the difference in infectivity for the two different cases.

memory capacity. We have shown that in addition to the pulse-type solution discussed at length in previous works, there are other radically different possible behaviors. Specifically, there can be an extended state in which the host population remains infected by a very broad range of viral strains. This state initially appears via a finite-amplitude instability of the pulse solution and leads to a pattern of endemic infection whose spread is only limited by the total host population and the ensuing demographic noise. We have studied this novel pattern both by direct simulation and also by a semianalytic approach relying on a cutoff mean-field deterministic version of the governing stochastic process. Our model contains, of course, many parameters, but our results suggest that a key determinant of the observed behavior is the range of immune coverage, here characterized by the ratio of the width w of the exponential falloff in immune efficacy to the phenotypic distance covered by viral evolution during a typical host infection time, here c/r where c is the pulse speed. Once the system reaches an extended configuration, its eventual size is determined by the effective cutoff, which here scales as $1/N_h$.

We have already discussed how adaptive systems with effective memory decay should generically exhibit the phenomenology described in this work. We wish to add that such a decay can occur at the individual level as in the variety of models considered above, each exploring a different specific mechanism, or could occur also at the population level. For example, imagine that we are considering populations with birth and death, the latter being necessary to address questions regarding disease virulence and its possible evolution. Then, new individuals would constitute a nonimmunized reservoir which could drive our tail instability. Another possibility is an infection that is spreading geographically, reaching previously uninfected hosts. Taken all together, we expect that many real-life situations could lead to the type of endemic states discussed here.

It is important to note that the tail instability provides a temporal pattern distinct from what would be seen in a system exhibiting the type of oscillatory state found by Sasaki and coworkers [13]. There, the pulse exhibits a "near-death" episode and only survives when small numbers of individuals at the leading edge start growing rapidly. In this scenario, a new strain would appear only after a period of time with no obvious infection; based on our preliminary investigations, this behavior is in fact very sensitive to the form of the cross-reactivity function and is not at all seen in models with the exponential decay used here (data not shown; results to be published). In our case, the basic infection continues to evolve continuously but spawns reinfection by strains that naively appeared to be vanquished.

There are many additional directions that could be studied in future work. The reduction of the problem to a one-dimensional phenotypic equation relies on the localization of the viral population in orthogonal dimensions, namely neutral genomic variations that do not contribute to fitness changes. This localization has been seen in simulations of the pulse solution ([11] and unpublished data) but has not to date been considered in the context of extended states. From the pure theory perspective, proving that the pulse instability is always finite-amplitude is a tractable problem. There is also the need to better understand the N_h dependence of the various patterns, possibly finding an explicit way to take the limit $N_h \to \infty$ for the pulse solution. It should also be possible to relate the "comb" solution to similar structures that can appear in the type of nonlocal Fisher equations used in the modeling of species diversity [17,18,31].

The real challenge for this line of research is making more quantitative contact with actual experimental data. This will require a detailed analysis of the cross-reactivity function. In the context of COVID-19, there is data and analysis regarding how antibodies effective against one strain lose effectiveness when confronted with other strains [32]. In the context of CRISPR-based phage immunity, there is a reasonable understanding of how changes in the viral sequence can alter the immune response [25]. Once this is accomplished, one should be able to look for the temporal signature of the tail instability. One might imagine comparing the spreading of the mound solution (before it reaches a steady state) with data on the dynamics of viral variants in a host population. A pulse solution corresponds to a replacement theory; a new strain takes over and no one gets infected by previous strains. A mound state on the other hand suggests the possibility that old variants can reappear even as old variants remain active. On the applied side, one could use our theoretical framework to design vaccine treatments that would interfere with the natural dynamics and increase the chances of collapse of the infected population.

ACKNOWLEDGMENTS

H.L. acknowledges the support of the National Science Foundation (NSF) of the Center for Theoretical Biological Physics, PHY-2019745. D.A.K. acknowledges the support of the Israel Science Foundation, 1614/21. We thank J. Levine for help with the schematic of the model.

APPENDIX A: PULSE SOLUTION: DETAILS

In this Appendix, we present more of the details of the argument arguing for at most a discrete set of propagation velocities for the isolated pulse solution, each with a specific number of infected hosts. As outlined in the main text, the essence of the argument is to analyze the number of adjustable parameters characterizing the solution of the deterministic equation in each of the two asymptotic regimes; the tail behind the front and the leading edge ahead of it.

Starting with the tail, we immediately conclude that the lack of new infections there means that the infected host density decays exponentially. Specifically, replacing $\dot{n}$ by $-cn'$ (i.e., going to the moving frame and assuming the pulse is moving to the right) and dropping I, we find from Eq. (2) that $n \approx n_b e^{rx/c}$ where n_b is an as yet unknown constant. Next we substitute this form into the definition of E in Eq. (13a); using the aforementioned form of the cutoff we arrive at

$$E \approx rR_0\epsilon^{-\alpha}e^{-(\alpha+1)rx/c}n_b^{\alpha+1}. \tag{A1}$$

From E we can uniquely determine V from Eq. (13b). Finally we note that ρ_m has a homogeneous solution and Q has a homogeneous term that can be added to the particular solution

driven by ρ_m,

$$\rho_m \approx \rho_{m,b} e^{\beta x} - \frac{n_b}{N_h M} e^{rx/c},$$
$$Q \approx \frac{2\rho_{m,b}}{1/w - w\beta^2} e^{\beta x} + Q_b e^{x/w}, \tag{A2}$$

where we have defined $\beta = N_I/(cMN_h)$. The inhomogeneous term is an exponentially suppressed relative to the inhomogeneous term, since $\beta < r/c$. Thus the most general solution in the tail has three unknown constants, n_b, $\rho_{M,b}$ and Q_b.

The situation is considerably more complicated at the leading edge, as here we have new infections occurring. The cutoff in Eq. (12a) plays a key role here, as mentioned above. Since the cutoff not only multiplies the E term in V, but also the "diffusion" term E'', we effectively have a case of nonlinear diffusion. Hence we might expect (and will verify below) that solutions exist for which the infected host population becomes precisely zero after some value of x. This is analogous to what has been seen in the Fisher equation with nonlinear diffusivity [22–24]; in particular, it has been shown there that the fronts that form from localized initial conditions have this specific property. Thus, we have to first find the specific form of this nonlinear solution and then determine, by linearizing away from that solution, the number of unknowns governing the fields in this region.

To accomplish this, we assume that

$$n(x) \approx n_f (1-x)^\gamma, \tag{A3}$$

where we have used translation invariance to fix the singular point to occur at $x = 1$. The left-hand side of the n evolution equation, Eq. (12a) is then $cn_f\gamma(1-x)^{\gamma-1}$. This term is larger than the $-rn$ term on the right-hand side and therefore it must be balanced by the infection piece I. As x approaches 1 from below, the leading term in I arises from the E'' term in V. Using the asymptotic form of the cutoff for small n, we have

$$E \approx rR_0\epsilon^{-\alpha} n^{\alpha+1} \approx rR_0\epsilon^{-\alpha} n_f^{\alpha+1}(1-x)^{(\alpha+1)\gamma},$$
$$V \approx \frac{\mu}{2\epsilon^\alpha} rR_0 n_f^{\alpha+1}(\alpha+1)\gamma[(\alpha+1)\gamma - 2](1-x)^{(\alpha+1)\gamma-2}. \tag{A4}$$

To get I, we just have to multiply by the as yet undetermined factor $\chi \equiv 1 - \frac{N_I}{N_h}$ and by $p_{\rm inf}(x) \approx p_{\rm inf}(1) = (1 - p_0 Q(1))^M$, since Q is determined globally and is not singular. Thus, the matching of the two sides of Eq. (12a) requires the powers of $(1-x)$ to be equal to each other, which yields $\gamma = 1/\alpha$. Then the coefficients must also match, fixing n_f to satisfy

$$n_f^{-\alpha} = \frac{\mu r R_0 \chi(1+\alpha)}{2c\epsilon^\alpha \alpha}(1 - p_0 Q(1))^M. \tag{A5}$$

This result shows that $Q(1)$ is an unknown number, adding to the list of parameters that need to be determined. Note that the integral formula for Q, Eq. (10) directly determines that $Q'(1) = -Q(1)/w$ and hence this is not an additional degree of freedom.

We now consider the linearized equation around this base solution so as to determine how many unknowns need to be specified in order to integrate the equations from the leading edge to the pulse center. Denoting the shifted values of n, E, and I as, respectively, $\delta n = A_n(1-x)^p$, $\delta E = A_E(1-x)^{p+1}$ and $\delta I = A_I(1-x)^{p-1}$, we have the following equations to leading order in $(1-x)$:

$$cpA_n = A_I, \quad A_I = (1 - p_0 Q(1))^M \frac{\mu}{2} p(p+1) A_E,$$
$$A_E = rR_0 n_f^\alpha (\alpha+1) A_n \epsilon^{-\alpha} = A_n \frac{2c\alpha}{\mu(1 - p_0 Q(1))^M}.$$

Combining these equations leads to the very simple condition, $p = p(p+1)\alpha$, which yields the two possibilities $p = 0$ and $p = 1/\alpha - 1$. The second of these just corresponds to a translation of the base solution (recall that $\gamma = 1/\alpha$) and neither of these are allowed. Thus there are *no* additional unknown constants in the leading edge solution. We also should mention that χ is an unknown parameter, but this will have to be determined self-consistently by the fact that it is directly related to N_I; picking any arbitrary value of χ will lead to this constraint being violated.

These results are used in the main text to show that the number of unknown parameters (including the velocity as one of these) equals the number of matching conditions. Thus, constructing a global solution will fix the velocity, leading to there being at most a discrete set of allowed pulses. In fact, we never find more than one such solution.

APPENDIX B: MOUND SOLUTION: DETAILS

In the main text, we considered the dimension of the solution space emerging when considering spatially dependent deviations from the constant bulk solution. We showed that there are two bad and two good modes, implying that the solution for large negative x has two unknown parameters, namely the coefficients of the two good modes.

We here turn to the front edge. Unlike what was the case for the propagating front, the time-independent equation for n is purely algebraic and hence n can be formally eliminated, leaving us with a two-field problem in terms of E and Q. In detail, we will eliminate n via Eq. (13a); since the right-hand side is a monotonically increasing function of n, we can define $g(x)$ such that

$$n(E) = g(E/rR_0), \tag{B1}$$

so that Eq. (13a) translates to

$$g(\zeta)(1 - e^{-(g(\zeta)/\epsilon)^\alpha}) = \zeta. \tag{B2}$$

Then the coupled differential equations read

$$E + \frac{\mu}{2}E'' = \frac{rn(E)}{(1 - p_0 Q(x))^M \chi},$$
$$Q'' - \frac{Q}{w^2} = -\frac{2}{w}\frac{n(E)}{N_I}.$$

We now have to count degrees of freedom, and to do this we have to investigate the behavior of E near $x = 1$. Writing $E \approx E_1(1-x)^\gamma$, we first need the small argument expansion of g:

$$g(\zeta)^{\alpha+1}/\epsilon^\alpha \approx \zeta \quad \Rightarrow \quad g(\zeta) \approx \epsilon^{\alpha/(\alpha+1)} \zeta^{1/(\alpha+1)}, \tag{B3}$$

so that

$$\frac{\mu}{2}\gamma(\gamma-1)E_1(1-x)^{\gamma-2} \approx r\epsilon^{\alpha/(\alpha+1)}\left(\frac{E_1(1-x)^{\gamma}}{rR_0}\right)^{1/(\alpha+1)}\frac{1}{(1-p_0Q(1))^M\chi}. \tag{B4}$$

From this, we can read off γ:

$$\gamma-2=\frac{\gamma}{\alpha+1}\Rightarrow\gamma=\frac{2(\alpha+1)}{\alpha}. \tag{B5}$$

This implies that $n\sim(1-x)^{\gamma-2}=(1-x)^{2/\alpha}$. Solving now for E_1 yields

$$E_1^{\alpha/(\alpha+1)}=\frac{2r}{\mu\gamma(\gamma-1)}\epsilon^{\alpha/(\alpha+1)}\left(\frac{1}{rR_0\chi}\right)^{1/(\alpha+1)}\times\frac{1}{\left(1-p^0_{imm}Q(1)\right)^{N_M}}, \tag{B6}$$

so that

$$E_1=\left(\frac{\alpha^2}{\mu(\alpha+1)(\alpha+2)(1-p_0Q(1))^M\chi}\right)^{(\alpha+1)/\alpha}\times r\epsilon\left(\frac{1}{R_0}\right)^{1/\alpha}. \tag{B7}$$

This implies that

$$n\approx\left(\frac{\alpha^2}{\mu(\alpha+1)(\alpha+2)(1-p_0Q(1))^MR_0}\right)^{1/\alpha}\epsilon(1-x)^{2/\alpha}. \tag{B8}$$

As for the pulse, we now need to investigate the behavior of the linearized equations about this nonlinear solution. Writing

$$E\approx E_1(1-x)^{\gamma}+\delta_E;\quad Q\approx Q(1)+\delta_Q, \tag{B9}$$

we have the linearized system [to leading order in $[1-x)]$

$$\begin{aligned}\frac{\mu}{2}\delta_E''&=\delta_E\frac{rn'(E)}{(1-p_0Q(1))^M}=r\delta_E\frac{n(E)}{E(1+\alpha)(1-p_0Q(1))^M}\\&\approx\delta_E\frac{\mu(E'')}{2E(1+\alpha)}=\delta_E\frac{\mu(\alpha+2)}{\alpha^2(1-x)^2},\\\delta_Q''&=-\frac{2}{w}\delta_E\frac{n'(E)}{N_I}=-\frac{2}{w}\frac{n(E)}{(1+\alpha)EN_I}\delta_E\\&=-\frac{\mu E'}{rwE(1+\alpha)N_I}\delta_E\approx-\frac{2\mu(\alpha+2)}{rw(1-x)^2N_I}\delta_E.\end{aligned} \tag{B10}$$

The assumption $\delta_E\sim(1-x)^p$ is consistent with the spatial dependence of these two equations. Substituting this into the first equation, we get the indicial equation for the exponent p,

$$p(p-1)=\frac{2(\alpha+2)}{\alpha^2}\Rightarrow p=\left\{\frac{\alpha+2}{\alpha},-2\alpha\right\}. \tag{B11}$$

The first of these choices, being $\gamma-1$, corresponds to translations (i.e., moving the location of the singular point), and so is ruled out. The second is singular and is also forbidden, so in fact the solution for E is unique. Q satisfies the boundary condition $Q'(1)=-(1/w)Q(1)$, so the only degree of freedom at the edge is $Q(1)$.

[1] T. Mora and A. M. Walczak, Quantitative theory of viral-immune coevolution may be within reach, PRX Life **1**, 011001 (2023).

[2] I. S. Novella, E. A. Duarte, S. F. Elena, A. Moya, E. Domingo, and J. J. Holland, Exponential increases of RNA virus fitness during large population transmissions., Proc. Natl. Acad. Sci. USA **92**, 5841 (1995).

[3] M. Worobey and E. C. Holmes, Evolutionary aspects of recombination in RNA viruses, J. Gen. Virol. **80**, 2535 (1999).

[4] T. B. Kepler and A. S. Perelson, Modeling and optimization of populations subject to time-dependent mutation, Proc. Natl. Acad. Sci. USA **92**, 8219 (1995).

[5] L. S. Tsimring, H. Levine, and D. A. Kessler, RNA virus evolution via a fitness-space model, Phys. Rev. Lett. **76**, 4440 (1996).

[6] I. M. Rouzine, J. Wakeley, and J. M. Coffin, The solitary wave of asexual evolution, Proc. Natl. Acad. Sci. USA **100**, 587 (2003).

[7] M. M. Desai, D. S. Fisher, and A. W. Murray, The speed of evolution and maintenance of variation in asexual populations, Curr. Biol. **17**, 385 (2007).

[8] J. A. G. De Visser and J. Krug, Empirical fitness landscapes and the predictability of evolution, Nat. Rev. Genet. **15**, 480 (2014).

[9] O. Hallatschek, The noisy edge of traveling waves, Proc. Natl. Acad. Sci. USA **108**, 1783 (2011).

[10] I. M. Rouzine and G. Rozhnova, Antigenic evolution of viruses in host populations, PLoS Pathog. **14**, e1007291 (2018).

[11] J. Marchi, M. Lässig, A. M. Walczak, and T. Mora, Antigenic waves of virus-immune coevolution, Proc. Natl. Acad. Sci. USA **118**, e2103398118 (2021).

[12] V. Chardès, A. Mazzolini, T. Mora, and A. M. Walczak, Evolutionary stability of antigenically escaping viruses, Proc. Natl. Acad. Sci. USA **120**, e2307712120 (2023).

[13] Y. Haraguchi and A. Sasaki, Evolutionary pattern of intra-host pathogen antigenic drift: Effect of cross-reactivity in immune response, Philos. Trans. R. Soc. London B **352**, 11 (1997).

[14] E. Cohen, D. A. Kessler, and H. Levine, Front propagation up a reaction rate gradient, Phys. Rev. E **72**, 066126 (2005).

[15] R. A. Neher and O. Hallatschek, Genealogies of rapidly adapting populations, Proc. Natl. Acad. Sci. USA **110**, 437 (2013).

[16] L. A. Segel and A. S. Perelson, Shape space: An approach to the evaluation of cross-reactivity effects, stability and controllability in the immune system, Immunol. Lett. **22**, 91 (1989).

[17] S. Pigolotti, C. López, and E. Hernández-García, Species clustering in competitive Lotka-Volterra models, Phys. Rev. Lett. **98**, 258101 (2007).

[18] T. Rogers, A. J. McKane, and A. G. Rossberg, Demographic noise can lead to the spontaneous formation of species, Europhys. Lett. **97**, 40008 (2012).

[19] E. Brener, H. Levine, and Y. Tu, Mean-field theory for diffusion-limited aggregation in low dimensions, Phys. Rev. Lett. **66**, 1978 (1991).

[20] E. Brunet and B. Derrida, Shift in the velocity of a front due to a cutoff, Phys. Rev. E **56**, 2597 (1997).

[21] L. Pechenik and H. Levine, Interfacial velocity corrections due to multiplicative noise, Phys. Rev. E **59**, 3893 (1999).

[22] W. I. Newman, Some exact solutions to a non-linear diffusion problem in population genetics and combustion, J. Theor. Biol. **85**, 325 (1980).

[23] F. Sánchez-Garduño and P. K. Maini, Existence and uniqueness of sharp traveling waves in degenerate non-linear diffusion Fisher-KPP equations, J. Math. Biol. **33**, 163 (1994).

[24] E. Ben-Jacob, I. Cohen, and H. Levine, Cooperative self-organization of microorganisms, Adv. Phys. **49**, 395 (2000).

[25] M. E. Bonomo and M. W. Deem, The physicist's guide to one of biotechnology's hottest new topics: CRISPR-Cas, Phys. Biol. **15**, 041002 (2018).

[26] S. C. Garrett, Pruning and tending immune memories: Spacer dynamics in the CRISPR array, Front. Microbiol. **12**, 664299 (2021).

[27] W.-C. Cao, W. Liu, P.-H. Zhang, F. Zhang, and J. H. Richardus, Disappearance of antibodies to SARS-associated coronavirus after recovery, N. Engl. J. Med. **357**, 1162 (2007).

[28] C. Willyard, Are repeat COVID infections dangerous? What the science says, Nature (London) **616**, 650 (2023).

[29] H. Chemaitelly, N. Nagelkerke, H. H. Ayoub, P. Coyle, P. Tang, H. M. Yassine, H. A. Al-Khatib, M. K. Smatti, M. R. Hasan, Z. Al-Kanaani *et al.*, Duration of immune protection of SARS-CoV-2 natural infection against reinfection, J.Travel Medicine **29**, taac109 (2022).

[30] O. N. Bjørnstad, K. Shea, M. Krzywinski, and N. Altman, The SEIRS model for infectious disease dynamics, Nat. Methods **17**, 557 (2020).

[31] Y. A. Ganan and D. A. Kessler, Front propagation and clustering in the stochastic nonlocal Fisher equation, Phys. Rev. E **97**, 042213 (2018).

[32] P. Padmanabhan, R. Desikan, and N. M. Dixit, Modeling how antibody responses may determine the efficacy of COVID-19 vaccines, Nat. Comput. Sci. **2**, 123 (2022).